

Online Appendix for Reducing Frictions in College Admissions: Evidence from the Common Application

Brian Knight and Nathan Schiff*

December 29, 2020

A Theoretical Model

A.1 Additional notation

There is a unit mass of type-1 students and a unit-mass of type-2 students. Colleges have a fixed capacity $\kappa \in (0, 1)$; thus, 2κ represents overall capacity across the two institutions.

We solve the model backwards. In the third stage, following student application decisions and university admissions offers, there are four possible student choice sets: choosing between both colleges, college 1 or 2 only, or only the outside option. The value of being accepted to both colleges equals $C_{12} = \ln[\exp(U_1) + \exp(U_2) + 1]$.¹ Likewise, we denote $Y_{12} = \exp(U_1)/[1 + \exp(U_1) + \exp(U_2)]$ as the yield for the first-choice college and y_{12} as the yield for the second-choice college for students accepted to both. The value of being accepted to only the first choice equals $C_1 = \ln[\exp(U_1) + 1]$, and the corresponding yield is $Y_1 = \exp(U_1)/[1 + \exp(U_1)]$. Similar expressions apply to the value of being accepted to only the second choice (C_2) and the corresponding yield is denoted by Y_2 , with $Y_1 > Y_2$.

In the second stage, taking yield as given, schools set their admission rates in order to satisfy capacity. We focus on an equilibrium in which all students apply to their first choice and a fraction b of students also apply to their second-choice college. Then, admissions rates are set in order to equate the number of student acceptances of university admissions offers to university capacity. For college 1, for example, total students acceptances equal the yield on first-choice students who are admitted to college 1 plus the yield on second-choice students who both apply to and are admitted to college 1. This must then equal the overall university capacity, as expressed below:

$$0.5Q_1 \underbrace{[(1-b)Y_1 + bQ_2Y_{12} + b(1-Q_2)Y_1]}_{\text{yield on first-choice}} + 0.5Q_1b \underbrace{[Q_2y_{12} + (1-Q_2)Y_2]}_{\text{yield on second-choice}} = \kappa \quad (1)$$

Among first-choice students, a fraction $1 - b$ apply to only their first choice, with yield of Y_1 , and a fraction b also apply to their second choice. In the latter case, a fraction Q_2 are also admitted to their second choice, with yield of Y_{12} , and a fraction $1 - Q_2$ are denied admission to their second choice,

*Knight: Brown University, 64 Waterman Street, Providence, RI 02906, brian.knight@brown.edu. Schiff: Shanghai University of Finance and Economics, 777 GuoDing Road, Shanghai, China, nschiff@gmail.com.

¹This follows the standard formula for consumer surplus in a logit model. Similar derivations apply for type 2 students, given the symmetry of the model.

with yield for college 1 thus equal to Y_1 . The second term represents yield on second-choice students, with a fraction b applying to both colleges. Among these, a fraction Q_2 are also admitted to their first choice and yield thus equals y_{12} . The remaining fraction $(1 - Q_2)$ are not admitted to their first choice and yield on these students equals Y_2 .

Then, in the first stage, applying to both colleges yields a value of $A_{12} = Q_1Q_2C_{12} + Q_1(1 - Q_2)C_1 + (1 - Q_1)Q_2C_2 - F - f$ for type 1 students. That is, students are accepted to both colleges with probability Q_1Q_2 , college 1 only with probability $Q_1(1 - Q_2)$, college 2 only with probability $(1 - Q_1)Q_2$ and face application costs of $F + f$. For type 1 students applying to only college 1, the value equals $A_1 = Q_1C_1 - F$. In equilibrium, the fraction of students applying to both colleges increases until the value from a second application equals the value of a single application ($A_{12} = A_1$). This can be written as:

$$Q_2[\underbrace{Q_1(C_{12} - C_1)}_{\text{option-value}} + \underbrace{(1 - Q_1)C_2}_{\text{safety-value}}] = f \quad (2)$$

The option value from a second application represents the benefit of being able to attend college 2 when the student has been accepted to both colleges, which occurs with probability Q_2Q_1 . This captures the idea that students may learn that college 2 is actually preferred to college 1 throughout the admissions process, following the realization of ϵ_1 and ϵ_2 . The safety value from a second application represents the benefit of being able to choose college 2 if not admitted to college 1, and this event occurs with probability $Q_2(1 - Q_1)$.

A.2 Conditions for an Interior Solution

Regarding equation 2, the key condition for a unique solution is that the upper solution to the quadratic equation implies an admissions rate in excess of one. To ensure that only the lower solution is feasible requires that application costs be small, relative to the benefits of a larger choice set:

$$F < C_{12} - C_1 \quad (3)$$

That is, the cost of a second application must be less than the option value of also being admitted to one's second choice. The requirement that F is small also guarantees that a solution exists, in the sense that the discriminant is positive.

Regarding equation 1, we require the following condition for an interior solution:

$$QY_1 < \kappa < Q^2(Y_{12} + y_{12}) + Q(1 - Q)(Y_1 + Y_2) \quad (4)$$

where Q is set at its equilibrium value and is thus a function of model parameters. The left hand side of the inequality requires that college capacity is more than sufficient to accommodate accepted students when all students apply to only their first choice, given equilibrium admissions rates. The right hand side requires that the college capacity is not sufficient to accommodate the situation when all students apply to both colleges, given equilibrium admissions rates. Thus, capacity can be neither too small nor too large.

A.3 Equilibrium Solution

We first solve equation 2 for the equilibrium admissions rate. While this equation is quadratic in Q and thus has two solutions in principle, the upper solution implies an admissions rate in excess of 1, under the assumptions outlined above, and we thus focus on the lower solution:

$$Q^* = \frac{C_2 - \sqrt{C_2^2 - 4f(C_1 + C_2 - C_{12})}}{2(C_1 + C_2 - C_{12})} \quad (5)$$

Given this equilibrium admissions rate, one can then calculate the equilibrium fraction of students applying to both colleges via equation 1, yielding:

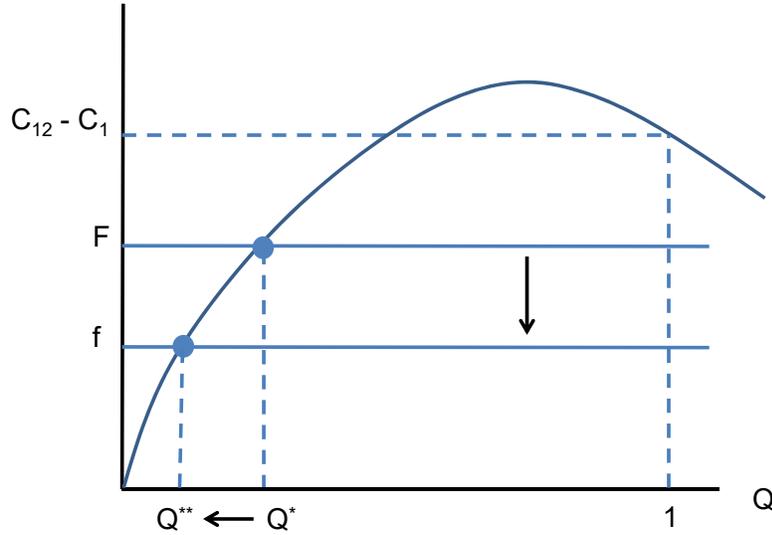
$$b^* = \frac{\kappa - QY_1}{Q^2[Y_{12} + y_{12} - Y_1] + Q(1 - Q)Y_2} \quad (6)$$

where Q is set at equilibrium levels.

A.4 Proof of Proposition 1

Parts 1) and 2): In Equation 5, it is clear that equilibrium admissions rates are increasing in F . Thus, a marginal reduction in F leads to a reduction in equilibrium admissions rates. This effect is illustrated in Figure A1 below.

Figure A1: Effects on admissions rates



Given that Q declines under the CA, we must next show that b is decreasing in Q . Taking the derivative of equation 6 with respect to Q , we have:

$$\frac{db}{dQ} = \frac{-Y_1}{D} - \frac{(\kappa - QY_1)[2Q(Y_{12} + y_{12} - Y_1 - Y_2) + Y_2]}{D^2} \quad (7)$$

where the denominator equals $D = Q^2[Y_{12} + y_{12} - Y_1] + Q(1 - Q)Y_2$. This denominator is positive since $Y_{12} + y_{12} > Y_1$.

Substituting back in the definition of b , we have that:

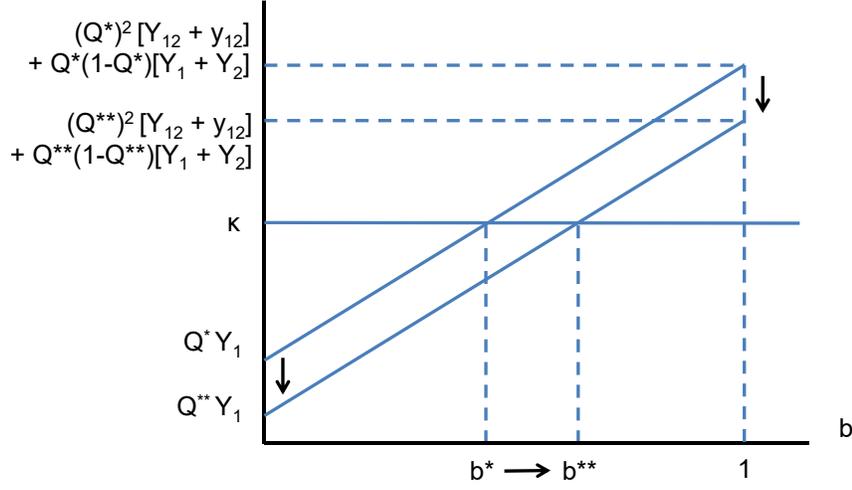
$$\frac{db}{dQ} = \frac{-Y_1 - b[2Q(Y_{12} + y_{12} - Y_1 - Y_2) + Y_2]}{D} \quad (8)$$

Re-arranging the numerator, this relationship can be written as follows:

$$\frac{db}{dQ} = \frac{-2bQ(Y_{12} + y_{12} - Y_1) + (bQY_2 - Y_1) - bY_2(1 - Q)}{D} \quad (9)$$

Each of these three terms in the numerator are negative. In particular, the first term is negative since $Y_{12} + y_{12} > Y_1$. The second term is negative since $Y_2 < Y_1$, $b < 1$, and $Q < 1$. Finally, the third term is negative since $Q < 1$ in equilibrium. Since the denominator must be positive for b to be positive, the slope is negative. This change in application rates is illustrated in Figure A2 below.

Figure A2: Effects on applications



Part 3): Note that the number of admitted students is equal to $Q(1 + b)$, the product of the admissions rate and the number of applications received. Using the closed form solution for b , This can be written as:

$$Q(1 + b) = Q + Qb = Q + \frac{\kappa - QY_1}{Q[Y_{12} + y_{12} - Y_1] + (1 - Q)Y_2} \quad (10)$$

Taking the derivative, we have that:

$$\frac{dQ(1 + b)}{dQ} = 1 - \frac{Y_1}{D} - \frac{\kappa - QY_1}{D^2}[Y_{12} + y_{12} - Y_1 - Y_2] \quad (11)$$

where the denominator equals $D = Q[Y_{12} + y_{12} - Y_1] + (1 - Q)Y_2$.

Using the fact that $Qb = [\kappa - QY_1]/D$, the slope can be re-written as:

$$\frac{dQ(1 + b)}{dQ} = 1 - \frac{Y_1}{D} - \frac{Qb}{D}[Y_{12} + y_{12} - Y_1 - Y_2] \quad (12)$$

This can be re-written as:

$$\frac{dQ(1 + b)}{dQ} = \frac{D - Y_1 - Qb[Y_{12} + y_{12} - Y_1 - Y_2]}{D} \quad (13)$$

Since D is positive, we simply need to show that the numerator is negative. Since the term $Y_{12} + y_{12} - Y_1 - Y_2$ is negative, the numerator is increasing in b . Thus, to show that it is negative

for all b between 0 and 1, we simply need to show that it is negative when $b = 1$. In this case, and canceling terms, the numerator can be written as $Y_2 - Y_1$, which is negative.

Part 4): The increase in out-of-state students follows directly from the increase in b resulting from the reduction in F .

A.5 Extension to 3 Colleges

We next consider the case in which two colleges ($c = 1$ and $c = 2$) join the CA but a third college ($c = 3$) does not join. In this case, there are three types of students, corresponding to the ex-ante ranking of the third college. Type 1 students have ex-ante preferences that rank college 3 last (there are two sub-types: either $U_1 > U_2 > U_3$ or $U_2 > U_1 > U_3$). Type 2 students have ex-ante preferences that rank college 3 in the middle (either $U_1 > U_3 > U_2$ or $U_2 > U_3 > U_1$). Type 3 students have ex-ante preferences that rank college 3 first (either $U_3 > U_1 > U_2$ or $U_3 > U_2 > U_1$). Given all of this, we can write the ex-ante preferences of the three different types (six different sub-types) of students as follows:

$$\begin{aligned}
 U_1 > U_2 > U_3, 1.1 \\
 U_2 > U_1 > U_3, 1.2 \\
 U_1 > U_3 > U_2, 2.1 \\
 U_2 > U_3 > U_1, 2.2 \\
 U_3 > U_1 > U_2, 3.1 \\
 U_3 > U_2 > U_1, 3.2
 \end{aligned}$$

Let Q_{CA} and Q_N denote admissions rates at the CA colleges and the non-CA college, respectively. Capacities are symmetric and equal κ .

We focus here on the case in which students do not apply to all three colleges.² Let b_1 be the fraction of type 1 students applying to their first and second choice and likewise for b_2 and b_3 . The capacity constraint for college 1 (college 2 is analogous) is now given by:

$$\begin{aligned}
 & Q_{CA}[\underbrace{(1 - b_1)Y_1}_{type\ 1.1} + \underbrace{(1 - b_2)Y_1}_{type\ 2.1}] + \\
 & \underbrace{b_1Q_{CA}Y_{12} + b_1(1 - Q_{CA})Y_1}_{type\ 1.1} + \underbrace{b_2Q_NY_{12} + b_2(1 - Q_N)Y_1}_{type\ 2.1} + \quad (14)
 \end{aligned}$$

$$\begin{aligned}
 & Q_{CA}[\underbrace{b_1Q_{CA}Y_{12} + b_1(1 - Q_{CA})Y_2}_{type\ 1.2} + \underbrace{b_3Q_NY_{12} + b_3(1 - Q_N)Y_2}_{type\ 3.1}] = \kappa \quad (15)
 \end{aligned}$$

And for college 3 it is:

$$\begin{aligned}
 & 2Q_N[\underbrace{(1 - b_3)Y_1 + b_3Q_{CA}Y_{12} + b_3(1 - Q_{CA})Y_1}_{types\ 3}] + \\
 & 2Q_N[\underbrace{b_2Q_{CA}Y_{12} + b_2(1 - Q_{CA})Y_2}_{types\ 2}] = \kappa \quad (16)
 \end{aligned}$$

²This could be formalized by having a large difference in preferences between the second and third choice. For type 1.1 students, for example, $U_2 - U_3$ would be large.

Prior to the CA, the three relevant indifference conditions are given, similarly to before, by:

$$Q^2(C_{12} - C_1) + (1 - Q)QC_2 = F \quad (17)$$

With the introduction of the CA, the conditions become as follows for type 1 students:

$$Q_{CA}^2(C_{12} - C_1) + (1 - Q_{CA})Q_{CA}C_2 = f \quad (18)$$

However, for types 2 and 3, now they incorporate the two different admission rates. For type 2 students, taking the case of 2.1, we have that:

$$Q_{CA}Q_N(C_{12} - C_1) + (1 - Q_{CA})Q_NC_2 = F \quad (19)$$

For type 3 students, taking the case of 3.1, we have that:

$$Q_NQ_{CA}(C_{12} - C_1) + (1 - Q_N)Q_{CA}C_1 = F \quad (20)$$

Claim: these three indifference conditions cannot be simultaneously satisfied

Proof: The introduction of the CA causes the right hand side in equation 17 to fall from F to f . This implies that, given the admissions rate at CA schools, more type 1 individuals find it profitable to apply to a second school, increasing b_{CA} . However, under a higher b_{CA} , CA colleges will have excess demand, violating their capacity constraint (equation 15). Thus, Q_{CA} must decrease until type 1 students are indifferent between applying to a second school or not.³

Now, note that the fall in Q_{CA} causes the left hand side in equation 19 to increase, implying that more type 2 students want to apply to a second college. However, the opposite happens with type 3 applicants. The fall in Q_{CA} pushes down the left hand side in equation 20, implying that fewer type 3 students want to apply to a second college.

Due to these opposing effects of a decrease in Q_{CA} for type 2 and type 3 students, both conditions cannot be simultaneously satisfied, meaning that either b_2 or b_3 must be at a corner solution. More formally, comparing the conditions for type 2 and type 3, we have that:

$$Q_{CA}Q_N(C_{1,2} - C_1) + (1 - Q_{CA})Q_NC_2 = Q_NQ_{CA}(C_{1,2} - C_1) + (1 - Q_N)Q_{CA}C_2$$

This is only satisfied when $Q_{CA} = Q_N$. However, under this condition, the left hand side of the three conditions are equal. But this is a contradiction with the fact that the first equation equals f , the second and third equations equal F , with $f < F$.

Claim: There is no equilibrium with $b_2 = 1$ and b_3 interior.

Proof: Assuming that b_1 is interior, and imposing symmetry, this would require the following:

$$Q_{CA}^2(C_{1,2} - C_1) + (1 - Q_{CA})Q_{CA}C_2 = f$$

$$Q_{CA}Q_N(C_{1,2} - C_1) + (1 - Q_{CA})Q_NC_2 > F$$

$$Q_NQ_{CA}(C_{1,2} - C_1) + (1 - Q_N)Q_{CA}C_2 = F$$

Comparing the conditions for type 1 and type 3 and using the fact that $f < F$, we have that:

³Note that an increase in b_{CA} must accompany the fall in Q_{CA} . Else, if only Q_{CA} were to fall, colleges would not meet the capacity constraint, as they would have open vacancies given the smaller admission rate.

$$Q_{CA}^2(C_{1,2} - C_1) + (1 - Q_{CA})Q_{CA}C_2 < Q_N Q_{CA}(C_{1,2} - C_1) + (1 - Q_N)Q_{CA}C_2$$

Re-arranging, this can be written as:

$$Q_{CA}(Q_{CA} - Q_N)(C_{1,2} - C_1 - C_2) < 0$$

Since $C_{1,2} - C_1 - C_2 < 0$, this requires $Q_{CA} > Q_N$.

Comparing types 2 and 3, we have that:

$$Q_{CA}Q_N(C_{1,2} - C_1) + (1 - Q_{CA})Q_N C_2 > Q_N Q_{CA}(C_{1,2} - C_1) + (1 - Q_N)Q_{CA}C_2$$

Re-arranging, this can be written as:

$$(1 - Q_{CA})Q_N > (1 - Q_N)Q_{CA} \tag{21}$$

This requires that $Q_N > Q_{CA}$. This contradicts that earlier requirement that $Q_{CA} > Q_N$.

Summary: The introduction of the CA leads to an increase in b_1 , the fraction applying to both CA schools and a reduction in Q_{CA} . Given this, it must be case that b_2 increases to 1 or that b_3 decreases to zero since both cannot be interior. However, we have shown that b_2 cannot equal 1, meaning that $b_3 = 0$. Thus, there is a reduction, all the way to zero, in the fraction applying to a school outside of the CA and a school inside the CA. Given this, there are network effects with more type 1.1 students attending college 2 and more type 1.2 students attending college 1. Likewise, there are fewer type 3.1 students attending college 1 and fewer type 3.2 students attending college 2.

Quantitative analysis: To provide further evidence on the three college case, we choose parameters that guarantee interior solutions before the policy change. In particular, we set $U_1 = 2$, $U_2 = 1$, $\kappa = 0.55$, and $F = 0.3$. Prior to the CA, colleges and students are symmetric, with $Q = 0.2985$ and $b = 0.0776$. Introduction of the CA lowers F to $f = 0.29$. Under the CA, imposing that $b_3 = 0$ in the new equilibrium, the admission rate of the CA schools falls from $Q = 0.2985$ to $Q'_{CA} = 0.2844$, while the admission rate of the non-CA school also falls, but by a smaller degree, to $Q'_N = 0.2942$. These changes reflect the direct and indirect effects of the decrease in F on the different types of students. Type 1 students are the only ones that benefit directly by the policy and strongly increase their applications to a second school from $b = 0.0776$ to $b'_1 = 0.3346$. On the other hand, type 2 students see their application rate grow only marginally, to $b'_2 = 0.0908$. The reason is that type 2 students have as first choice a CA college followed by the non-CA college, so they do not enjoy the lower application fee but do face the lower admission rate from the CA school, providing them with incentives to apply to their second choice.

A.6 Extension to Test Scores

We consider three colleges (c) and two test score types: low and high. We assume that colleges want to attract as many high test score students as possible and thus admit them with probability one. Low test score students are then admitted at a lower rate in order to fill any remaining capacity. Given our interest in stratification, we can then simply study the behavior of high test score students. Given that high test score students are admitted with certainty, the model plays out differently in this case. In particular, students will not be indifferent when choosing their application sets, and corner solutions are now relevant for these high test score students. Given these corner solutions, we focus on non-marginal changes in application costs.

We focus here on two cases. In the first case, application costs are sufficiently high that high test score students only apply to their first choice in the absence of the CA. For students with the preference order $U_1 > U_2 > U_3$, this requires:

$$C_{12} - C_1 < F \tag{22}$$

Suppose now that colleges 1 and 2, but not college 3, join the CA. Further, suppose that application costs fall sufficiently such that $C_{12} - C_1 > f$, and likewise for students with preference ordering $U_2 > U_1 > U_3$. Then, these two sets of students will apply to both colleges, and students with other preference orderings are unaffected. Since these two sets of students are now more likely to attend college (recall that $Y_{12} + y_{12} > Y_1$), the fraction of high test score students at CA colleges increases. The fraction of high test score students at colleges outside of the CA is unchanged.

In the second case, suppose that application costs are sufficiently low that high test score students apply to their top two choices, but not their third choice, in the absence of the CA. For students with the preference order $U_1 > U_3 > U_2$, not applying to the third college requires:

$$C_{123} - C_{13} < F \tag{23}$$

where C_{123} represents the value from having a full choice set of all three colleges. Suppose now that colleges 1 and 2, but not college 3, join the CA, and application costs fall sufficiently such that $C_{123} - C_{13} > f$, and likewise for all students that have college 1 or 2 as their third choice. Then, all students except those with preference orderings $U_1 > U_2 > U_3$ and $U_2 > U_1 > U_3$ will apply to all three colleges. Thus, there is an increase in applications for colleges 1 and 2 and no increase in applications for college 3. Given that the yield on students accepted to college 3 now falls (resulting from more college 3 applicants also applying to colleges 1 and 2), this implies that colleges 1 and 2 will now draw some high test score students who would have attended college 3 in the absence of the CA. Thus, as in the first case, the fraction of high test score students at CA colleges increases. The new effect here is that the fraction of high test score students falls at schools outside of the CA.

A.7 Extension to Student Income

We consider three colleges (c) and two income types: low-income and high-income. There are two types of application costs. As before, the time cost of applying to a first CA college equals F and the time cost of applying to a second CA college equals $f \leq F$. The financial costs of applying to a first CA colleges equals ϕ and the cost of applying to a second CA college also equals ϕ . We simplify the model by assuming that low-income students can only apply to one college, perhaps due to credit constraints. We also assume that colleges do not distinguish between low-income and high-income students in terms of admissions probabilities. Given all of this, only high-income students decide whether or not to apply to a second college. Thus, all of the results from the first extension apply to high-income students but not low-income students. In particular, colleges 1 and 2, which are members of the CA, experience an increase in applications from high-income students, relative to college 3, which is not a member of the CA. Given this, colleges 1 and 2 ultimately enroll more high-income students, relative to college 3, which ends up attracting more low-income students.

Table B1: Summary Statistics

	mean	sd	min	max	count
Applications	4462.01	6469.93	29.00	97121.00	34519
Yield	0.41	0.17	0.03	1.00	34361
Admits	2672.44	3380.56	25.00	36088.00	34556
Enrollment	931.38	1130.04	25.00	11807.00	38359
Selectivity	0.70	0.18	0.04	1.00	34468
Fraction Out of State	0.31	0.25	0.00	1.00	37628
SAT 25th pctile	967.22	136.67	410.00	1510.00	28534
SAT 75th pctile	1184.60	130.36	720.00	1600.00	28540
Non-White Enroll %	0.32	0.23	0.00	1.00	27517
PhD Faculty	222.80	348.97	0.00	3792.00	33893
Application Fee	29.82	16.84	0.00	150.00	40881
Ugrad Enroll % Pell	0.43	0.22	0.00	1.00	26821

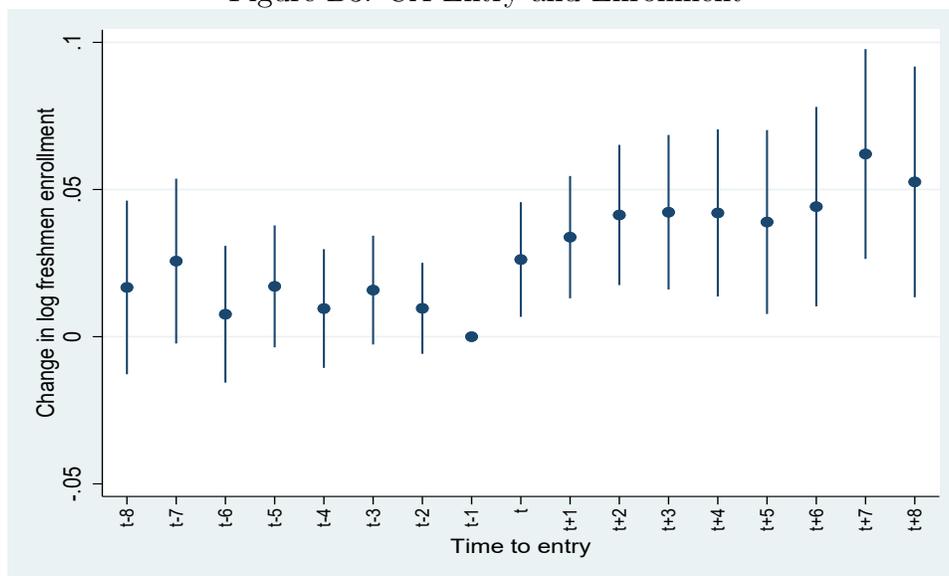
All variables from College Board data except Ugrad Enroll % Pell; this variable uses separate IPEDS and Dept. of Ed. data.

B Additional Empirical Results

B.1 Summary Statistics for Main Variables

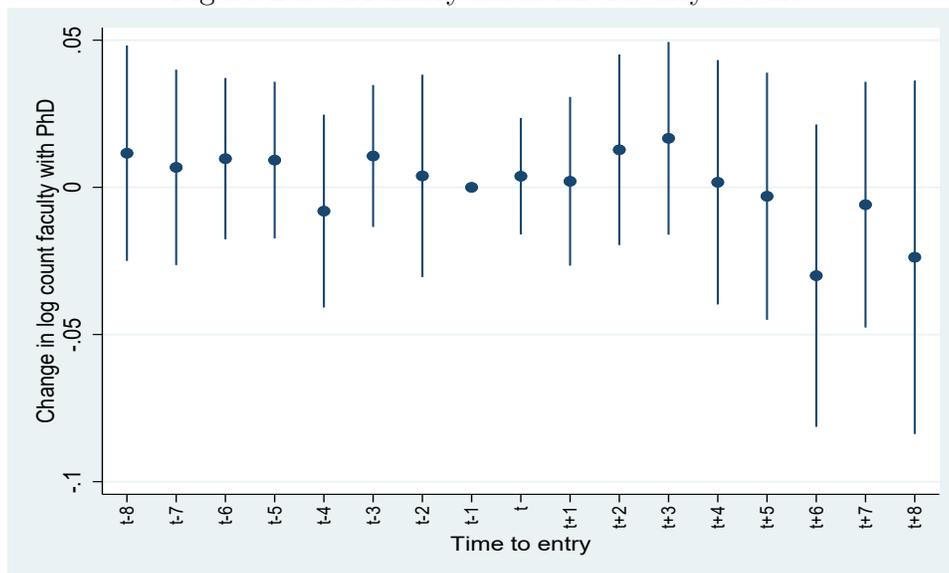
B.2 Event Study Plots for Additional Outcome Variables

Figure B3: CA Entry and Enrollment



Notes: Coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4 in main text) using the full sample.

Figure B4: CA Entry and PhD Faculty Count



Notes: Coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4 in main text) using the full sample.

B.3 Trends in Geographic Integration

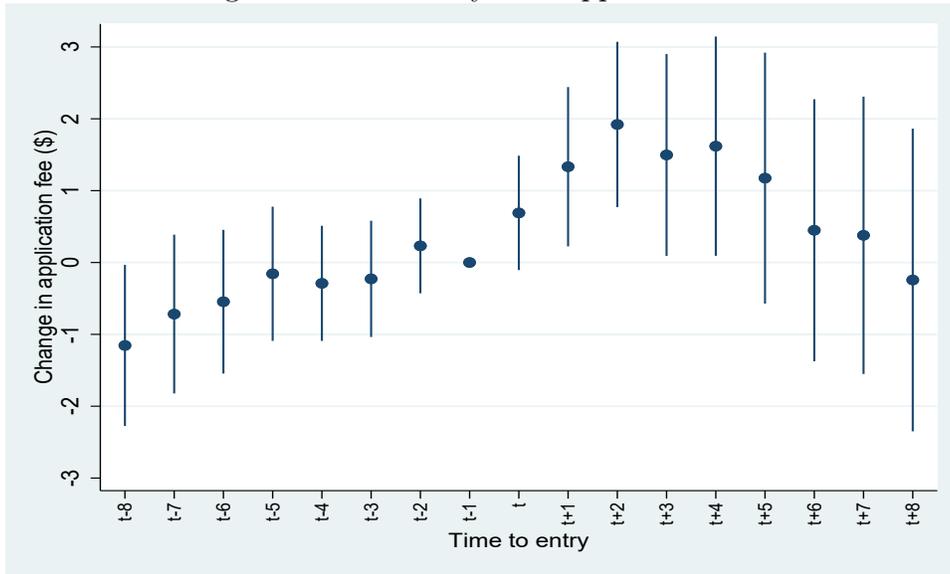
Hoxby (2000) documents that the percentage of students attending in-state institutions fell consistently from 1949 to 1994 and that the role of distance in explaining college choice decreased as well. We extend this study of geographic integration into our sample period by measuring trends in distance traveled from a student’s home state to the state of their university using IPEDS data over the period from 1986 to 2014.⁴ In Figure B6 we plot the mean distance traveled in each year, along with 95 percent confidence intervals.⁵ As shown, there is a clear increase in distance traveled over this time period, with the average distance traveled increasing by over 100 kilometers for private universities and roughly 40 kilometers for public universities. Thus, the trends towards greater geographic integration documented by Hoxby between 1949 and 1994 also appear over our sample period 1986-2014.

In Table B2 we calculate the average increase in geographic integration over time for public and private institutions, using the specification $y_{ct} = \beta_1 years_t + \beta_2 years_t \times public_c + \mu_c + \epsilon_{ct}$. In column 1 we find that distance traveled increases by about 3 kilometers per year for private institutions and 1 kilometer per for public institutions, while column 2 specifies average distance in logs and shows that both types of institutions have roughly the same percentage increase over time of 1.4 percent. In columns 3 and 4 we examine the average distance traveled by out-of-state students only, which allows us to distinguish the effect of a change in the percentage of out-of-state students from a change in the geographic composition of the out-of-state students. Interestingly, the results show that while

⁴In particular, we measure the great circle distance in kilometers between state centroids, defining the distance for all in-state students as zero

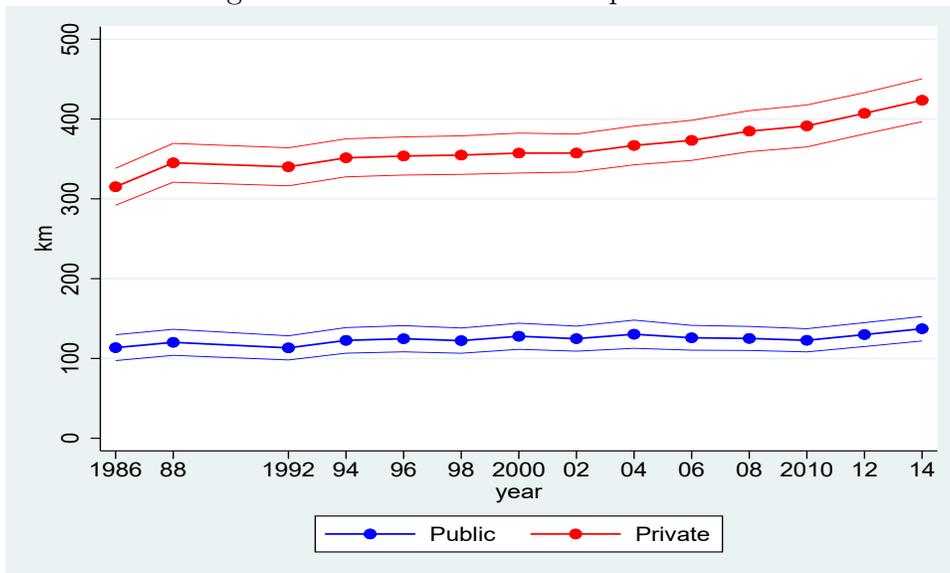
⁵Letting the subscript s denote a student’s state, we define the mean distance traveled by all students from US states at college c in year t as $avdist_{c,t} = (1/nat_enroll_{c,t}) \sum_{s \in S} enroll_{c,s,t} * dist_{c,s}$. The variable nat_enroll is total enrollment from the 50 US states and D.C. The home location of foreign students and students from US territories is usually not available, and therefore we excluded these groups from the total. However, students from these groups are counted in total enrollment when calculating percentage of students attending in-state.

Figure B5: CA Entry and Application Fee



Notes: Coefficients and 95% confidence intervals on the relative join year variables from the event study specification (Eq. 4 in main text) using the full sample.

Figure B6: Distance Traveled per Student



Notes: Plot shows mean distance traveled and 95% confidence intervals. Distance calculated between university-state centroid and student-state centroid. Distance for in-state students set to zero. Data covers every two years from 1986 to 2014, except 1990 (missing).

Table B2: Geographic Integration by Institution Type

	(1)	(2)	(3)	(4)	(5)
	Distance	Log Distance	Distance	Log Distance	Out-of-state %
years	3.1867*** (0.2797)	0.0136*** (0.0010)	5.1360*** (0.3278)	0.0070*** (0.0004)	0.0014*** (0.0002)
years X public	-1.9582*** (0.3571)	0.0001 (0.0018)	-5.5110*** (0.6052)	-0.0056*** (0.0007)	-0.0001 (0.0003)
Constant	236.9366*** (2.9090)	4.6789*** (0.0127)	876.4421*** (4.2712)	6.5852*** (0.0048)	0.2941*** (0.0022)
Observations	20685	20236	20236	20236	20685
Clusters	1708	1688	1688	1688	1708

Notes: Results from estimating $y_{ct} = \beta_1 years_t + \beta_2 years_t \times public_c + \mu_c + \epsilon_{ct}$, where μ_c is an institution fixed effect. The dependent variable in columns 1 and 2 is distance per student, in columns 3 and 4 it is distance per out-of-state student, while column 5 shows percentage of out-of-state students at the institution. Distance is measured in kilometers. Standard errors clustered by institution in parentheses.

out-of-state students at private institutions are traveling further each year, there is essentially no increase in distance for out-of-state students at public institutions (the interaction effect is the same magnitude as the main effect). This implies that the increasing distance traveled by public university students comes entirely from an increase in the out-of-state percentage, which increases at about 0.14 percentage points each year for both types of institutions.

B.4 CA and Geographic Integration (IPEDS data)

As shown in Table B3, universities, after joining the CA, experience a significant increase in the average distance students travel to attend, with column 1 documenting an increase of 30 kilometers, a roughly 10 percent increase (column 2). Restricting to only out-of-state students, the distance increases by 55 kilometers (column 3), an increase of 7 percent for this population (column 4). In addition to out-of-state students traveling further, joining the CA also decreases the fraction of in-state students by about 2.3 percentage points (column 5). Generally, the magnitudes of the effects in Table B3 are large. Comparing each coefficient in Table B3 to its counterpart in Table B2, the effect of joining is about 10 times larger than the yearly trend for distance measures and about 15 times larger for in-state percentage.

As a further analysis of the effect of joining CA on distance traveled, we now consider the change in the entire distribution of a college's enrollees over distance. To do so, we restrict our sample to only those institutions which joined the CA and which have migration data both three or four years before and three or four years after joining, depending on whether the institution joined in an odd or even year. We then sum the enrollees across all universities in each period (pre, post) and calculate the percentage coming from each state-to-state distance. This allows us to calculate two cumulative distribution functions (CDF), where each bin of the CDF represents a given state-to-state distance and we are calculating the percentage of students traveling that distance to all universities in a period. We then plot the difference between the before and after CDF in Figure B7.⁶ The largest difference

⁶We truncate the graph at 4000km since the share of students coming from a greater distance is very small in both periods.

Table B3: CA Entry and Geographic Integration

	(1)	(2)	(3)	(4)	(5)
	Distance	Log Distance	Distance	Log Distance	Out-of-state %
CA member	30.1654*** (6.1947)	0.1044*** (0.0255)	55.1233*** (8.5535)	0.0700*** (0.0105)	0.0233*** (0.0048)
Observations	20629	20176	20176	20176	20629
Clusters	1652	1628	1628	1628	1652

Notes: Results from constant coefficient specification (Eq. 3 in main text) on full sample, using IPEDS data. The dependent variable in columns 1 and 2 is distance per student, in columns 3 and 4 it is distance per out-of-state student, while column 5 shows percentage of out-of-state students at the institution. Distance is measured in kilometers. All specifications include institution and year fixed effects, standard errors clustered by institution in parentheses.

occurs at zero, indicating that most of the effect comes from a 3 percentage point decrease in the in-state percentage. The slope of this differenced CDF increases sharply and approaches zero, so that a distance of 1200 kilometers the change is less than 1 percent, and then flattens. This shape suggests that CA entry increases the distance traveled by enrollees by mostly increasing the number of enrollees from nearby states.

B.5 Future Joiners Comparison: Methodology and Data Preparation

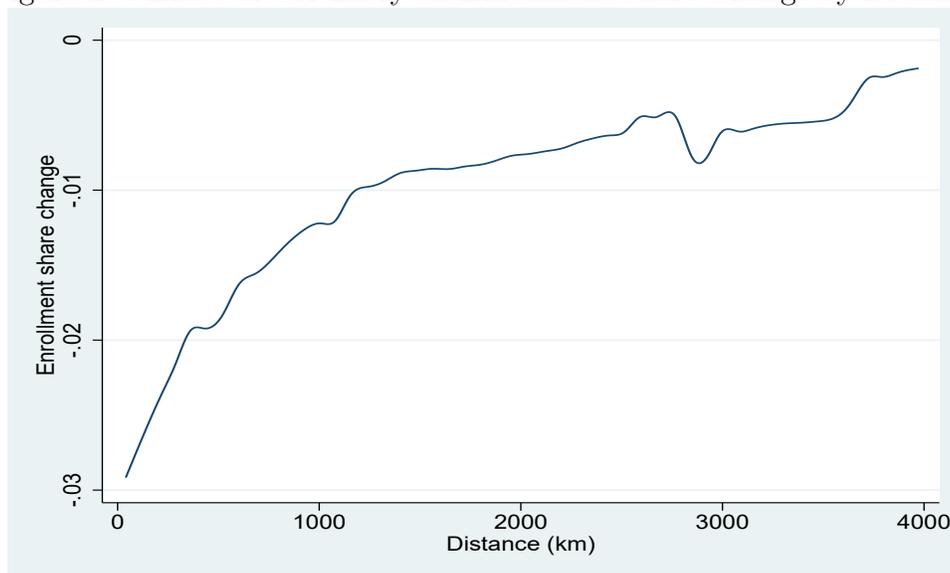
As described earlier, a key threat to identification in our analysis is that joiners might have different pre-trends relative to the comparison group, which includes both schools that never join during our sample period and schools that will join in the future but before the end of our sample period. We address this concern with an additional specification in which we compare outcomes for joiners to outcomes for colleges that will join the CA in the near future. Specifically, we compare schools that joined in year j to schools that will join in years $[j + 3, j + 5]$, over a pre-join period $[j - 5, j - 1]$ and a post-join period $[j, j + 2]$. This duration is short enough to make joiners and future joiners quite comparable, but still long enough to estimate a post-join effect. For example, for a school joining in 2000, the comparison group includes colleges that join in 2003, 2004, and 2005, and we analyze outcomes over the 1995-2002 period.

This empirical strategy is similar, although not identical, to that used by Deshpande and Li (2019) and we constructed our dataset in a similar way. We first assembled separate datasets of joiners and future joiners for every join year j , and then appended each join year’s data into a single dataset. The resulting dataset has some duplicate observations since the same school-year may serve as a comparison observation for multiple join-years. For example, a school joining in 2004 is a comparison observation for schools joining from 1999-2001. In all specifications we cluster standard errors at the school level and therefore these duplicates do not affect inference. Additionally, since we compare joiners to future joiners, most schools serve as both treated and comparison observations, over different join years j . Therefore, following Deshpande and Li (2019), we also include an additional indicator for whether a school is a joiner for a specific join year j as a comparison. The regression model using this strategy is:

$$\ln(y_{jct}) = \beta * (CA_{cjt}) + \mu_c + \mu_t + \alpha * 1(J_c = j) + \epsilon_{jct} \quad (24)$$

The corresponding event-study specification is:

Figure B7: Effect of CA Entry on Enrollment Share Change by Distance



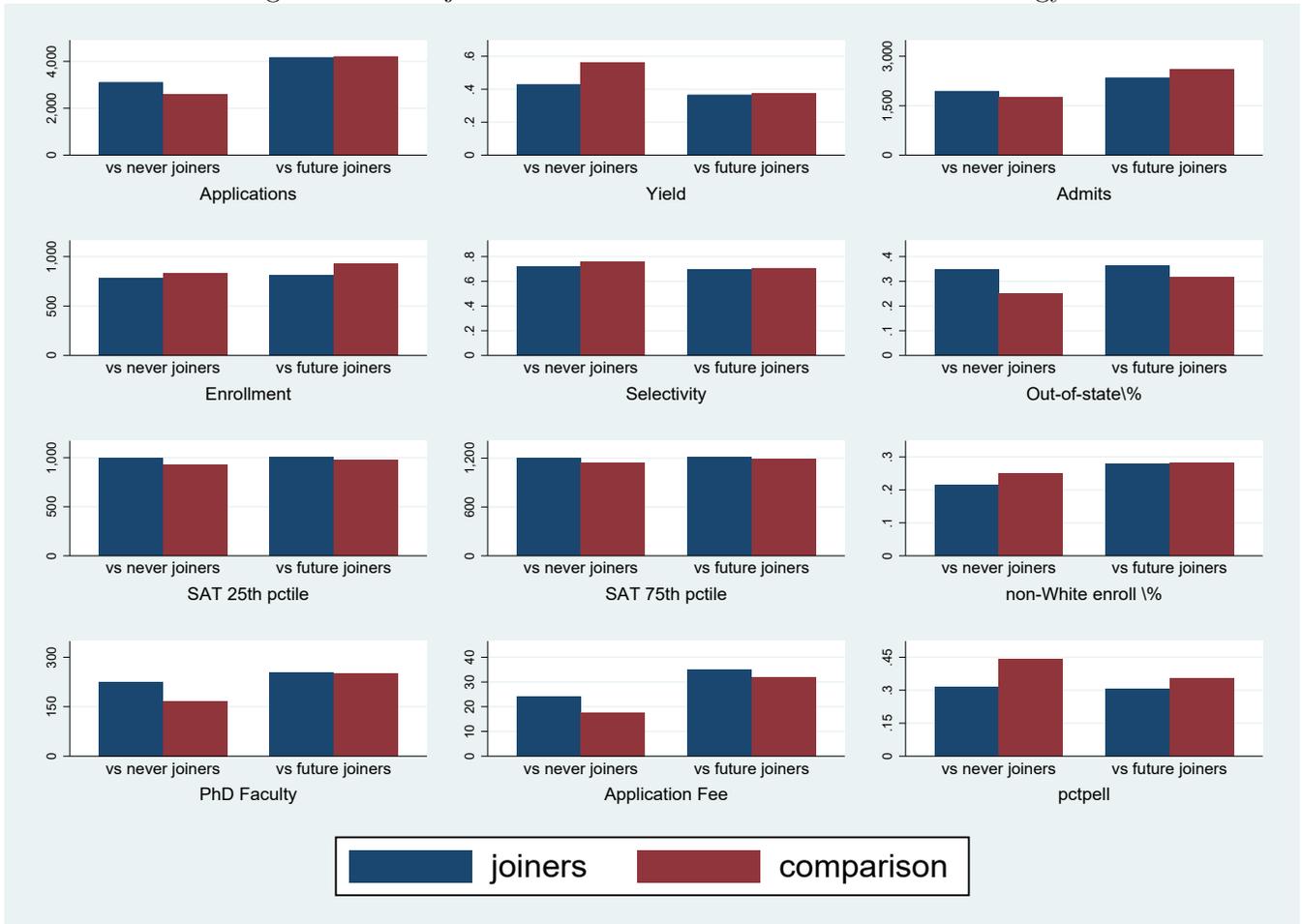
Notes: Enrollment share change defined as enrollment share before joining the CA minus the enrollment share after joining the CA. The before period is 3 to 4 years before the join year; the after period is 3 to 4 years after the join year. Distance is calculated between state centroids; in-state distance defined as zero. The graph is smoothed with the median spline method using 50 bands. Sample has 265 unique institutions.

$$\ln(y_{jct}) = \sum_{k=-w}^{w-1} \beta_k [1(J_c = j) \times 1(t - J_c = k)] + \mu_c + \mu_t + \alpha * 1(J_c = j) + \epsilon_{jct} \quad (25)$$

References

- Deshpande, Manasi, and Yue Li.** 2019. “Who is screened out? Application costs and the targeting of disability programs.” *American Economic Journal: Economic Policy*, 11(4): 213–48.
- Hoxby, Caroline M.** 2000. “The effects of geographic integration and increasing competition in the market for college education.”

Figure B8: Pre-join Differences for each Identification Strategy



Notes: First set of bars compares joiners to never joiners using 1990 values (1999 for Pell%). Second set of bars compares joiners to future joiners, using values from five periods before joining.